



METHODS AND SYSTEMS FOR AUTOMATICALLY CREATING A DATA FEED FILE FOR USE WITH DESKTOP APPLICATIONS

Cross-reference to Related Applications

This application is a continuation of United States Patent Application Serial No. 09/753,306 filed December 30, 2000, the contents of which are incorporated by reference herein in their entirety.

FIELD OF INVENTION

[0001] The present invention relates generally to methods and systems for data file creation. Embodiments of the present invention provide for automatic data extraction and data feed file creation at remote locations and transferring completed data feed files to locations electronically accessible to a software application.

BACKGROUND

[0002] The ability to gather and manipulate data is increasingly important in today's society. Small businesses and large corporations alike are aware that compiling, manipulating and reviewing data can provide valuable information about their present and future customers. Whether it is data on product usage or data on seasonal trends in customer purchases, the need to locate, gather on demand, and manipulate data is important. Even computer software designers are aware that the ability to manipulate and present data may provide valuable insight to business owners and managers. To that end, many software designers have developed software applications capable of manipulating and presenting data in a variety of ways to help improve data compilation and manipulation.

[0003] Companies have long recognized the need for gathering and storing data relating to its customers, suppliers etc. At one time, this data was kept in boxes and stored in rooms. When storage space for data became cumbersome, data was often transferred to mainframe computers and other like electronic data storage devices. These storage devices allow massive electronic data files to be stored in one location. However, when mainframes and other like devices became the preferred method of storing large volumes of data, the need

and the technology to perform data manipulations to present data in a variety of ways was not as prevalent as it is today. Instead, the mainframe was often used to archive data that was important but infrequently accessed.

[0004] As times changed so did the purpose of storing data on a mainframe and the frequency with which this data was accessed. No longer is data stored on a mainframe primarily because it is historic information that needs to be archived. To the contrary, data is sometimes stored on a mainframe because it is so voluminous that it would occupy too much space on a local area network server, a wide area network server or a desktop computer hard drive. The increased frequency with which data on a mainframes is needed creates problems with efficiently providing desktop software applications access to the data when it is stored on a remote mainframe or other like data storage device.

[0005] To obtain information stored in data files on a mainframe, end-users often have to print the entire data file from the mainframe. Oftentimes, the data manipulation performed by the desktop software application does not require all the data stored in the data file on the mainframe. The end-user is, therefore, forced to manually filter through the voluminous data file, pinpoint the data needed and manually create a data feed file for use by the desktop software application. Thus, there is an obvious need for the capability to automatically and electronically select the data stored on the mainframe and generate a data feed file that is accessible to desktop software applications for data manipulation and presentation.

SUMMARY OF INVENTION

[0006] Generally speaking, the present invention relates to a method and system for generating a data file by extracting specified data from a larger raw data file. The goal of this invention is to generate a data feed file that can be used by desktop applications for any number of operations including, but not limited to, data compilation and data tracking.

[0007] According to an embodiment of this invention, the raw data file contains large volumes of related data. The raw data file is comprised of multiple records, each record comprised of at least one data field. Each data field contains the same category of data in each record of the raw data file. Assume, for this example and all examples hereafter, that

the raw data file contains customer-billing information. Each record in the raw data file would therefore contain customer-billing information pertaining to an individual customer. Each record in the raw data file would further be comprised of data fields, each containing a different category of data, such as customer name, customer address, the customer phone number etc. Below is an example of the records and data fields of a raw data file.

	Data field (1)	Data field (2)	Data field (3)	Data field (4)
	Customer Name	Customer Address	Customer Phone Number
Record 1	Anne Joseph	123 Courtyard Dr	(555) 555-5555
Record 2	George Tilly	324 Patent Dr	(555) 444-1234

[0008] In an embodiment of this invention, the raw data file is so voluminous that it is stored on a remote data storage device, such as a mainframe, instead of being stored on a computer hard drive, local area network (LAN) or other like electronic storage device accessible to a user-preferred desktop software application or applications.

[0009] In an embodiment of this invention, desktop software applications include, but are not limited to, the following commonly know software applications: MICROSOFT EXCEL, MICROSOFT ACCESS, and ORACLE. These software applications and other like applications, are used to perform operations such as, data compiling, data tracking and other types of data manipulation. In an embodiment of this invention, the user-preferred software application uses data contained in a raw data file that resides on a mainframe or other like data storage device. In an aspect of the invention, the operation performed by the desktop application, may not require all the data contained in the raw data file. Instead, the desired data manipulation may only require select data contained in the raw data file. It is the goal of this invention to generate a data feed file by extracting only needed data from the raw data file. The data feed file will then contain only the data the user-preferred desktop software application needs to perform its designated operation.

[0010] According to an embodiment of this invention, the data feed file is generated by extracting the desired data from the raw data file and populating the data feed file with the

extracted data. To determine what data to include in the data feed file, the end-user develops a set of data feed file criteria.

[0011] In an embodiment of the invention, the data feed file criteria identifies the location of the desired data in the raw data file by identifying the data field in each record of the raw data file that contains the desired data. The data field containing the desired data may be identified in any number of ways. In another embodiment of the invention, the location of the data field is identified by the category of data the data field contains. For example, a data field in each record of the raw data file may contain the amount payable by each customer. In such cases, the data field could be identified as the "amount payable data field".

[0012] In another embodiment of the present invention, the location of the data field is identified by the position of the data field in relation to other data fields in each record of the raw data file. For example, the data field that contains the desired data may be the fifth data field from the left in each record of the raw data file. In such cases, the data field could be identified as data field number five (5).

[0013] In addition to identifying the location of the desired data in the raw data file, the data feed file criteria may specify additional requirements for data to be included in the data feed file. For example, the data feed file criteria may specify that data located in the amount payable data field should be included in the data feed file only if the amount payable is greater than two hundred dollars (\$200).

[0014] In yet another embodiment of the invention, the set of data feed file criteria may further contain criteria for including data not in any data field of the raw data file. For example, different service fees may apply to customers depending on the customer's zip code. However, the raw data file may not contain the various fees. In such cases, the data feed file criteria may specify that a data field in the data feed file include a designated service fee for customers depending on the customer's zip code.

[0015] In an embodiment of the invention, multiple raw data files may contain the data needed for the data feed file. In such cases, the data feed file criteria may further identify the raw data file from which specified data should be extracted.

[0016] Once the end-user develops a set of data feed file criteria for the data feed file, the end-user provides a data extraction tool the data-locating criteria. The data extraction tool then locates and extracts the desired data in the raw data file using the data-locating criteria. The data extraction tool further creates a report containing the extracted data.

[0017] Further, in another embodiment of the invention, data extracted from multiple data fields of the same record are placed on the same line in the report. The data extracted for each data field is separated by a space in the report. Each line of the report is also delineated by a print stop character.

[0018] As described above, the data feed file criteria may include criteria that defines the characteristics of data to include in the data feed file ("data-limiting criteria"). To ensure that the data feed file contains only the data specified by the data-limiting criteria, a discrimination tool reads the report created by the data extraction tool and verifies that the data on each line of the report complies with the data-limiting criteria. When the discrimination tool locates data on a line in the report that does not comply with data-limiting criteria, the discrimination tool deletes the entire line of data from the report. By eliminating the entire line of data, the discrimination tool eliminates all the unnecessary data from the report and prevents unnecessary information from being included in the data feed file.

[0019] In an embodiment of the present invention, the discrimination tool may not always be required. When the end-user only develops data-locating criteria, the discrimination tool may not be necessary and the step involving the discrimination tool may be eliminated.

[0020] After the report is created and includes only the data necessary to generate the data feed file, a formatting tool is used to complete the data feed file. The data feed file, like the raw data file, contains at least one record. Each record in the data feed file will further comprise at least one data field and each data field in each record will comprise a single category of data. The number of data fields in the data feed file is determined by the data feed file criteria. For example, if the data-locating criteria identifies the location of five (5) data fields in the raw data file from which data should be extracted, the data feed file will contain at least five data fields. Additional data fields may also be included in the data feed

file if the data feed file criteria designates data to be included in the data feed file that is not contained in the raw data file.

[0021] In an embodiment of the invention, the formatting tool uses the data from the report to populate each data field in each record of the newly created data feed file. The formatting tool may further populate fields in the data feed file with additional data specified by the data feed file criteria. The formatting tool may also arrange data fields in the data feed file in any order specified by the end user.

[0022] Once the formatting tool has populated the data fields of the data feed file, the data feed file is ready to be used by the user-preferred software application or applications. The completed data feed file, is thereafter transmitted through the appropriate network to a location electronically accessible to the user-preferred desktop software application. In one embodiment of the present invention, the data feed file is transferred using a file transfer protocol (FTP) to a local area network (LAN) accessible to the user-preferred desktop software application.

[0023] An advantage of this invention is that the data feed file is created automatically. The invention eliminates the need to manually review the raw data file and manually enter the desired data from a raw data file to use with the preferred desktop application.

[0024] It is a further object of the invention to improve the integrity of information generated by desktop applications by eliminating data entry errors that inevitably occur when a data feed file is manually generated.

[0025] It is a further object of this invention to reduce the time needed to compile data for use with desktop applications.

[0026] It is yet another object of this invention to improve the ability to change and update reports created by desktop applications using data located on mainframes or other electronic storage devices.

DESCRIPTION OF FIGURES

Figure 1 is a block diagram that shows the elements of a system used to carry out the steps of the present invention.

Figure 2 is a flow diagram illustrating a method of operation of the present invention.

Figure 3 is a flow diagram illustrating the method of operation of the data extraction tool.

Figure 4 is a flow diagram illustrating the method of operation of the discrimination tool.

Figure 5 is a flow diagram illustrating the method of operation of the formatting tool.

DETAILED DESCRIPTION

[0027] The embodiments included herein are intended in all respects to be illustrative rather than restrictive. Alternative embodiments will become apparent to those skilled in the art to which the present invention pertains without departing from its spirit and scope.

Overview of the Invention

[0028] The present invention comprises a method and system for automatically generating a data feed file for use with desktop applications. The method comprises: developing a set of data feed file criteria comprising at least one data-locating criterion, extracting at least a portion of data from a raw data file based on at least one data-locating criterion, creating a report wherein the report comprises the extracted data, and populating the data feed file with data from the report.

[0029] In an embodiment of the invention, the method further comprises transmitting the data feed file to a location electronically accessible to at least one desktop application.

[0030] In yet another embodiment of the invention, a local area network is the location electronically accessible to at least one software application. In addition, the hard drive of a desktop computer may be the location electronically accessible to at least one software application.

[0031] In an embodiment of the invention, the raw data file comprises at least one record, each record comprises at least one data field. Data in the raw data file is separated into the data fields of the raw data file.

[0032] In an embodiment of the invention, the set of data feed file criteria comprises data-locating criterion; data-limiting criterion, and hard-coded data criterion. The data-locating criterion identifies the location of the desired data in the raw data file. The data-limiting criterion identifies characteristics of the desired data. The hard-coded data criterion identifies data to include in the data feed file that is not located in the raw data file.

[0033] In an embodiment of the invention, the location of desired data in the raw data file is defined as the location of the data field containing the desired data.

[0034] In yet another embodiment of the invention, the data field location is identified by the category of data the data field contains. The data field location may also be identified by the position of the data field in the raw data file.

[0035] In an embodiment of the invention, lines of the report created using data extracted from the raw data file are deleted when data on the line does not comply with data-limiting criteria.

[0036] Another embodiment of the invention comprises populating the data feed file with data provided by the hard-coded data criteria.

[0037] Another aspect of the invention comprises a system for generating a data feed file from a raw data file for use by a software application the system comprising: at least two data storage devices; a data extraction tool wherein the data extraction tool extracts data from the raw data file and creates a report containing the extracted data; and a formatting tool wherein the formatting tool populates the data feed file with data extracted by the data execution tool.

[0038] In yet another embodiment of the invention, the system comprises at one two data storage devices accessible to at least one desktop application. More preferable, the data storage device accessible to at least one software application is a local area network or a

computer hard drive. Even more preferable, the data storage device that stores the raw data file is a mainframe.

[0039] In an embodiment of the invention, the system further comprises a means to electronically transmit data between at least two data storage devices.

[0040] In an embodiment of the invention, the data extraction tool extracts data from the raw data file at a locations in the raw data file identified by at least one data-locating criterion.

[0041] In an embodiment of the invention, the system further comprises a discrimination tool wherein the discrimination tool analyzes the report created by the data extraction tool and deletes lines from the report that contain data that does not comply with at least one data-limiting criterion.

[0042] In an embodiment of the invention, the formatting tool populates said data feed file with data from the report created by the data extraction tool. In another embodiment of the invention, the formatting tool further populates the data feed file with data identified by at least one hard-coded data criterion.

[0043] Further details of the method and system of the present invention are described below with reference to the appended figures.

Overview of an Embodiment of the Invention

[0044] Figure 1 is a block diagram that shows the elements of a system used to carry out the steps of the present invention. The embodiment comprises a raw data file 10 stored on a mainframe 24 or other electronic data storage device. In an embodiment of the invention, the raw data file 10 comprises large amounts of data separated into records. Each record in the raw data file 10 comprises at least one data field. More preferable, each record in the raw data file contains multiple data fields. Data in a raw data file is separated into data fields by the category of data contained in the data field. One aspect of the invention comprises, multiple raw data files 10 stored on a mainframe 24 or other data storage device.

[0045] A data extraction tool 14 is also shown in Figure 1. The data extraction tool 14 is used to locate and extract data from pre-identified locations in the raw data file 10. The data extraction tool 14 is further capable of creating a report containing the extracted data. In an embodiment of the present invention, the data extraction tool 14 is capable of locating and extracting data from multiple raw data files. The data extraction tool 14 is further capable of compiling, into one report, data extracted from multiple raw data files.

[0046] In an embodiment of the present invention, the data extraction tool 14 is a mainframe utility commonly known as UNITECH. While UNITECH is conventionally used as a printing and reconciliation tool designed to perform accounting operations, UNITECH may be customized to locate and extract data from pre-identified data fields in a raw data file 10 and create a report comprising the extracted data. Although an embodiment of the invention uses UNITECH as the data extraction tool 14, any tool capable of locating and extracting data from pre-identified locations and creating a report containing the extracted data can serve as a suitable data extraction tool 14.

[0047] A discrimination tool 12 is also shown in Figure 1. The discrimination tool 12 is used to locate and delete data from the report created by the data extraction tool 14 based on a set of data feed file criteria. In an embodiment of the invention, the discrimination tool 12, can delete entire lines of data from the report based on a set of criteria.

[0048] The discrimination tool 12, in an embodiment of this invention, is a mainframe utility commonly known as Record Select. While Record Select is the discrimination tool 12 used in the embodiment shown, any tool or software capable of locating and deleting entire lines of data from a report based on a set of criteria can serve as the discrimination tool 12.

[0049] In an embodiment of the invention, the discrimination tool 12 is not always needed. When data feed file criteria is limited to data-locating criterion, the step involving the discrimination tool 12 may be eliminated.

[0050] The formatting tool 16 shown in Figure 1 creates the data feed file template, extracts data from the report created by the data extraction tool 14 and populates the data feed file with the extracted data and any other data specified by the end-user. In an

embodiment of the invention, the completed data feed file is electronically transmitted through an appropriate network 20 to an electronic data storage device accessible to the user-preferred desktop application 22. The formatting tool 16, in an embodiment of this invention, is a mainframe utility commonly known as SyncSort. While SyncSort is the formatting tool 16 used in the embodiment shown, any tool or software capable of locating and deleting entire lines of data from a report based on a set of criteria can serve as the formatting tool 16.

[0051] In an embodiment of the present invention, the completed data feed file is electronically transmitted from the mainframe device 24 using a file transfer protocol (FTP) to a local area network (LAN) 22 accessible to the desired desktop application.

Overview of Method of Operation

[0052] A method of operation of the present invention can be described using Figures 2-5. Figure 2 is a flow diagram illustrating a method of operation of the present invention. In one embodiment of the invention, the end-user develops a set of data feed file criteria 30. The set of data feed file criteria 30 comprises information necessary to locate the desired data in the raw data file ("date-locating criteria").

[0053] In an embodiment of the invention, the data-locating criteria identifies the location of data fields in the raw data file that contain the desired data. The location of data fields containing desired data may be determined in a variety of ways. In an embodiment of the invention, the location of data fields is determined by identifying the category of data contained in the data field. For example, the data field that contains customer account number data may be identified as the customer account number data field.

[0054] In another embodiment of the invention, the location of the data fields containing desired data may be determined by identifying the position of the data field in the raw data file in relation to other data fields. For example, if the desired data is contained in the data field that is the fifth data field from the left in each record of the raw data file, the data field could be identified as data field number five (5).

[0055] Further methods of locating data in a raw data file will be apparent to one skilled in the relevant art.

[0056] In addition to data-locating criteria, the set of data feed file criteria may further include information to define the characteristics of the desired data ("data-limiting criteria"). For example, in addition to identifying the location of the data field that contains amounts payable data in a raw data file, the data-limiting criteria may also indicate the need for amounts payable data only if it is greater than two hundred dollars (\$200).

[0057] The data feed file criteria may further comprise criteria that specifies data to be included in the data feed file that is not located in the raw data file ("hard-coded data criteria"). For example, the data feed file criteria may specify that a data field should be included in the data feed file for service fees, the amount of which varies depending on the customer's zip code. The service fees are not, however, contained in the raw data file. In such cases, the hard-coded data criteria specifies the appropriate fee for each zip code to be included in the generated data feed file.

[0058] The set of data feed file criteria 30 provides instruction for the data extraction tool 14, the discrimination tool 12, and the formatting tool 16, all of which are shown in Figure 1. Once the set of data feed file criteria is developed 30, the data extraction is activated 32 and the data-locating criteria is provided to the data extraction tool by the end-user or some other similarly situated individual. The method of operation of the data extraction tool 14 is illustrated in Figure 3.

[0059] Figure 3 shows that the data extraction tool receives data-locating criteria. The data extraction tool then uses the data-locating criteria to locate data in the raw data file 50. Most preferably the data extraction tool locates data in the raw data file 50 based on the location of the data field containing the desired data. When the data extraction tool locates the desired data in the raw data file 50, the data extraction tool extracts the located data from the raw data file 52.

[0060] After the data extraction tool extracts the data from the raw data file 52, the data extraction tool creates a report 54. The report is comprised of all the data that the data

extraction tool extracted from the raw data file. All data extracted from a single record in the raw data file is placed on the same line in the report. Data extracted from each data field is separated in the report by a space and each line in the report ends with a print control character. The print control character serves as a signal that the data extracted from a single record in the raw data file is included on that line in the report.

[0061] The report created by the data extraction tool contains all the data extracted from the raw data file based on data-locating criteria. However, as mentioned above, the data feed file criteria may also specify data-limiting criteria. When the data feed file criteria includes data-limiting criteria, the report created by the data extraction tool must be analyzed by the discrimination tool for compliance with data-limiting criteria. Thus, as shown in Figure 2, if data-limiting criteria is developed 33 the discrimination tool is activated 34 and provided the data-limiting criteria after the data extraction tool extracts data from the raw data file and creates the report. On the other hand, if data-limiting criteria is not developed, the formatting tool is activated 36 after the data extraction tool creates the report containing extracted data.

[0062] Figure 4 illustrates the operation of the discrimination tool. To analyze the report for compliance with data feed file criteria, the discrimination tool is provided the data-limiting criteria 59. The discrimination tool then reads each line in the report 60 to determine whether data in the report complies with the data-limiting criteria. If data on any line in the report does not comply with the data-limiting criteria 62, the discrimination tool deletes the entire line from the report 64.

[0063] For example, assume that the data-limiting criteria indicated that customer-billing data for the data feed file should only include data regarding customers whose amount payable is greater than two hundred dollars (\$200). In this case, the data extraction tool would locate and extract the amount payable, regardless of amount, for each customer in the raw data file. The amount payable would then be placed in the report created by the data extraction tool. The discrimination tool then reads the report to locate lines of data on which the amount payable is less than two hundred dollars (\$200). The discrimination tool would then delete each line in the report that contains an amount payable less than two hundred dollars (\$200).

[0064] It is possible that the discrimination tool is not utilized in every application of this invention. For example, it is possible that the data feed file criteria will only provide data-locating criteria. In such cases, the step involving the discrimination tool may be disregarded.

[0065] When the report created by the data extraction tool contains only the data that will be included in the data feed file, the formatting tool is triggered as shown in Figure 2. Figure 5 illustrates the method of operation of the formatting tool. The formatting tool 16 creates a template of the data feed file 70 using the data feed file criteria. For example, if the data feed file criteria indicates that the desired data is located in five (5) data fields of the raw data file, the formatting tool creates a data feed file template that contains at least one record comprising at least five (5) data fields. The number of additional data fields will be determined by the hard-coded data criteria. The formatting tool then reads the report 72. The formatting tool further extracts data from the report and populates the data fields 74 in the data feed file with the data from the report.

[0066] In an embodiment of the invention, the formatting tool reads the report 70 line by line from left to right. The formatting tool reads data from left to right in the report until it reaches a space in the report. When the formatting tool reaches a space in the report, the formatting tool populates the data field in the raw data file 74 with data located just before the space. The formatting tool continues this process until it reaches the end of the line in the report. In an embodiment of the present invention, the formatting tool also populates the data feed file as specified by the hard-coded criteria. For example, if the hard-coded criteria states that records with a designated zip code should also include a service fee of \$2.00, the formatting tool locates records in the data feed file that meets that criteria and populates the data field in the data feed file with the appropriate service fee.

[0067] In an embodiment of the invention, the end of a line in the report is signaled by a print stop character located at the end of each line in the report. When the formatting tool reaches the print stop character at the end of the line in the report, the formatting tool starts to populate the next record in the data feed file by reading the next line in the report from left to right and populating data fields 74 in the data feed file in the matter described above. The

process of extracting data from the report and populating data fields in a record of the data feed file continues until all the data in the report is put in the data feed file, i.e. until the data extraction tool reaches the end of the report 76.

[0068] Returning to Figure 2, after the formatting tool has populated the data feed file, the data feed file is ready to be transferred to a location accessible to the user-preferred desktop application 38. In an embodiment of the invention, the data feed file is transferred, using the appropriate network, to a location accessible to the preferred desktop application for which the data feed file was created. The location accessible to the preferred desktop application may be a desktop computer's hard drive, a local or wide area network or any other like location electronically accessible to the preferred desktop application. The method of electronically transferring a data file from one location to another will be apparent to one skilled in the relevant art.